

## Research Article

# Microbial biodiversity and metabolic functioning in sediments of coastal dune lakes on a remote island

John K. Pearman<sup>1</sup>, Jack Sissons<sup>1</sup>, Joseph Kanyi Kihika<sup>1</sup>, Georgia Thomson-Laing<sup>1</sup>, Susanna A. Wood<sup>1</sup>

<sup>1</sup> Molecular and Algal Ecology Group, Cawthron Institute, Nelson, New Zealand

Corresponding author: John K. Pearman ([john.pearman@cawthron.org.nz](mailto:john.pearman@cawthron.org.nz))

## Abstract

Molecular-based techniques provide the potential for novel insights into the functioning of ecosystems, especially those that are globally rare such as coastal dune lakes. In the surface sediments of lakes, microbial communities play a vital role in biogeochemical cycling and techniques such as metagenomics can provide information on the roles play in these ecosystems. The current study aimed to investigate the taxonomic and functional composition of six coastal dune lakes on Chatham Island using sediment DNA approaches. The use of metabarcoding (16S rRNA gene) and metagenomics showed that there were distinct differences in the microbial community composition and functional potential amongst the lakes, especially in the lakes with higher salinity. Investigation of metabolic potential with metagenomics showed that the abundance of genes involved in nitrogen cycling were related to the nitrogen:phosphorus ratio while assimilatory sulfate reduction was correlated with sulfur and organic matter concentrations. Analysis showed differences in the carbon fixation strategies amongst the lakes. The lake with the highest salinity levels also had elevated levels of osmoprotectants and related transporters. The sequencing of sediment DNA enables the investigation of the composition and functioning of lake environments providing a basis for the increased understanding of the processes occurring within lakes.

**Key words:** Bacteria, dune lakes, lake sediment, metabarcoding, metagenomics, osmoprotection



Academic editor: Alexander Probst

Received: 8 January 2025

Accepted: 9 April 2025

Published: 27 May 2025

**Citation:** Pearman JK, Sissons J, Kihika JK, Thomson-Laing G, Wood SA (2025) Microbial biodiversity and metabolic functioning in sediments of coastal dune lakes on a remote island. *Metabarcoding and Metagenomics* 9: e144128. <https://doi.org/10.3897/mbmg.9.144128>

**Copyright:** © John K. Pearman et al.  
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

## Introduction

Lakes are inherently connected to their surrounding catchments, and their sediments act as sinks where nutrients and environmental contaminants accumulate (Li et al. 2024). Microbes are present in high quantities in the surface sediments and are important in many key processes in lakes including biogeochemical cycling. Microbial composition and function vary between lakes and this has often been attributed to differences in environmental conditions (Zhou and Ning 2017; Biessy et al. 2022). A variety of environmental factors can impact the microbial community including abiotic factors within the lake (e.g., pH, nutrient concentrations and temperature; Ruuskanen et al. 2018; Pearman et al. 2020) as well as effects derived from the wider environment including land use (Kraemer et al. 2020), altitude (Li et al. 2017; Pearman et al. 2020) and climate zones (Li et al. 2024).

Biological interactions (e.g. predation and competition) have also been shown to affect lake microbial communities (Zhou and Ning 2017). Enhancing knowledge of the factors that drive changes in microbial communities and their functioning is vital for predicting how lakes will respond to future environmental pressures.

A growing number of studies investigating microbial communities are moving beyond metabarcoding and the assessment of compositional changes. They are now focusing on metagenomics and the investigation of the functional potential of the microbes (Grossart et al. 2020). Many of these studies are linking microbial dynamics with biogeochemical processes. Metagenomics studies within aquatic systems have highlighted the importance of microbial communities and their functions in carbon fixation (Alfreider et al. 2017; Linz et al. 2018; Fang et al. 2022), as well as nitrogen and sulfate cycling (Biessy et al. 2022; Fang et al. 2022; Liu et al. 2022; Pearman et al. 2022b). Metagenomic research on lakes has often focused on the impact of nutrients on the lake communities (Biessy et al. 2022; Kuang et al. 2023). Total nitrogen and total phosphorus have been shown to affect microbial communities in lakes (Peura et al. 2015, Kuang et al. 2023) but the ratio of the two nutrients is also of interest as it can indicate if biological communities are being limited by one of the nutrients or the other (Downing and McCauley 1992). Salinity has also been shown to affect the functional potential of lake microbial communities with microbes having various methods to cope with osmotic challenges (Wood 2015). These can include the production of osmolytes (e.g., glycine betaine and proline) or the use of various transporters to respond to osmotic challenges (Wood 2015). Recent work, using metagenomes, has shown that across salinity gradients the functional potential of pathways including in nitrogen and sulfur cycling are affected by changes in osmolarity (Liu et al. 2022). Indeed, changes in the functional potential of microbes revealed by metagenomics have been suggested as a way to aid decision-making for bioremediation (Nandy and Kapley 2024). Despite the increasing research on lake microbiomes, globally rare systems, such as coastal dune lakes, remain underexplored especially in remote and isolated locations such as the Chatham Islands.

In this study, we used metagenomics to investigate microbial communities and their functional potential in six coastal dune lakes on a remote island (Chatham Island) which is part of Rēkohu (Chatham Islands) in Aotearoa New Zealand. Coastal dune lakes are rare globally. They are permanent lakes formed inland of dune systems and are restricted to areas of the United States, Mexico, Madagascar, eastern Australia and Aotearoa New Zealand (VanTassel and Janosik 2019). Within Aotearoa New Zealand most of the dune lakes are situated within Northland on the North Island or on Rēkohu. Rēkohu lies in the Pacific Ocean about 800 km east of Aotearoa New Zealand and consists of 10 islands, the largest of which is Chatham Island towards the eastern end of the Chatham Rise. The islands emerged within the last three million years due to tectonic uplift and the young age of the islands alongside the geographic isolation, variable climate and geology have combined to produce a distinctive set of ecosystems (Meredith and Croucher 2007). The anthropogenic nutrient input into the lakes has been low, and combined with the natural enrichment of phosphate (Meredith and Croucher 2007) the lakes of Rēkohu are potentially nitrogen limited. While freshwater systems on Chatham Island have received little attention, surveys have shown lakes to be depurated in eukaryotic macro-organism biodiversity and free from introduced European or North American exotic plants or fish that dominate lakes in

mainland Aotearoa New Zealand (Champion and Clayton 2004). Human arrival in the Rēkohu is believed to have occurred around 1500 CE (McFadgen 1994) with population densities on Rēkohu being low. The anthropogenic impacts on the lakes have been limited, with specific contamination discharges rare, although there has been substantial land use change over the centuries with deforestation occurring and land being converted to pasture (McFadgen 1994).

This research aimed to describe the microbial diversity and the functional profiles of six coastal dune lakes that are free from introduced exotic plants and fish. We hypothesized that; (1) despite relatively low anthropogenic influences, nitrogen and phosphorus would affect the community composition and functional potential of the lake sediment communities, and (2) that some of the microbes in the lakes with high salinities would have unique functional profiles.

## Methods

### Study Lakes and sampling

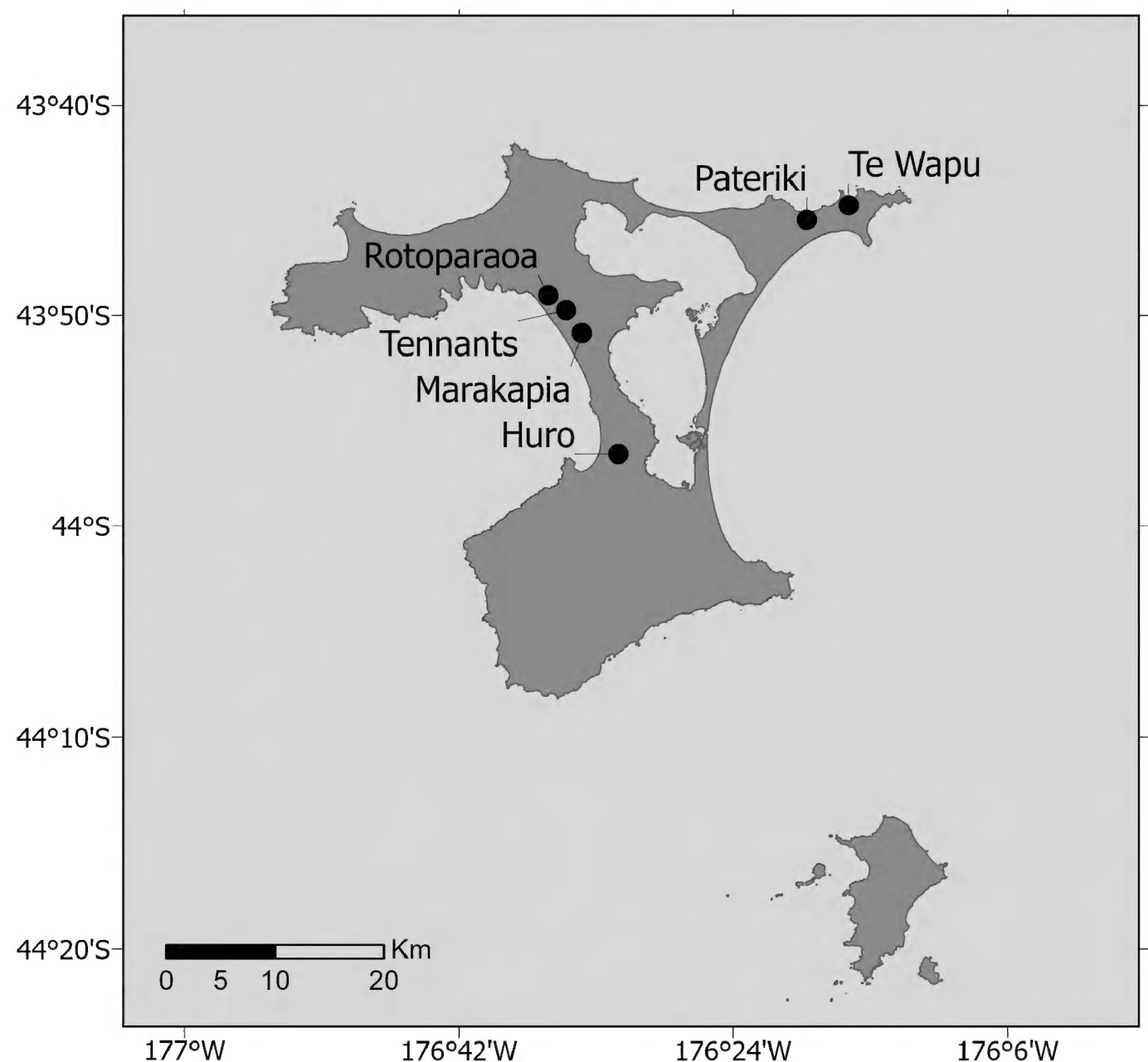
Six dune lakes on Chatham Island were sampled in April 2021 (Fig. 1; Suppl. material 1). The lakes ranged in depth from 1 m (Marakapia) to 6 m (Tennants). Lake Huro was the largest lake (588 ha) followed by Pateriki (134 ha) with the smallest lake being Rotoparaoa (5 ha). Four lakes were positioned on the west coast while the other two (Te Wapu and Pateriki) were on the north-east coast. These latter two had elevated salinity at the time of sampling due to intermittent connectivity with the ocean (Meredith and Croucher 2007).

Surface water samples (1 L) were collected from the lake's deepest point which was located using a depth sounder (Hawkeye H22PX, USA). Approximately 200 mL of water was filtered through a single GF/C filter and the volume was recorded. Filters were stored in the dark (-20 °C) before analysis for chlorophyll *a* (chl-*a*) was undertaken following the APHA 10 200H method at Watercare Laboratories (Auckland, Aotearoa New Zealand). The reporting limit was 0.0006 mg/L.

At each lake, ponar grabs (2.4L) were used to collect triplicate surface sediment samples from the deepest part of the lake. Using sterile spatulas, approximately 4 g of undisturbed surface sediment (1–2 mm depth) was placed in sterile tubes and frozen for later metabarcoding analysis. The remaining top 2 cm of sediment was collected in 500 mL containers and shipped chilled to the laboratory for analysis. Subsamples (~20 g) were taken for metagenomic analysis and frozen (-20 °C) until processing with the remaining sediment being used for nutrient and elemental analysis.

### Nutrient and elemental sampling and processing

The sediment for nutrient and elemental analysis was homogenized, centrifuged and pore water removed before elements were analyzed on an Inductively Coupled Plasma-Mass Spectrometer using the methods described in Pearman et al. (2020). The elements assessed were iron (Fe), manganese (Mn), aluminum (Al), calcium (Ca), lead (Pb), copper (Cu), zinc (Zn), cadmium (Cd), phosphorus (P) and sulfur (S) with reporting limits of (mg/kg) 12.5, 0.125, 2.5, 12.5, 0.05, 0.075, 0.05, 0.005, 10 and 250, respectively. Total nitrogen and total



**Figure 1.** Map of the sampling sites on Chatham Island.

organic carbon were analyzed using catalytic combustion while organic matter was determined by oven drying and gravimetric determination as described in Pearman et al. (2020).

### Molecular processing

All molecular work was undertaken in dedicated UV-sterilized laboratories under sterile conditions using laminar flow cabinets with HEPA filtration.

For the metabarcoding, DNA was extracted from 18 (six lakes in triplicate) surface sediment samples as detailed in Pearman et al. (2022a). Briefly, 0.25 g wet weight of sediment was processed using the DNeasy PowerSoil Kit (Qiagen, USA) following the manufacturer's instructions using a QIAcube extraction robot (Qiagen, USA). A negative extraction control (nuclease-free water) was also processed every 23<sup>rd</sup> sample. The V3–V4 region of the 16S rRNA gene was amplified with PCR using the 341F and 805R primers (Herlemann et al. 2007, Klindworth et al. 2013) and sequenced on an Illumina MiSeq platform (paired-end; 2 × 300 bp) at the Auckland Genomics Center (Auckland, Aotearoa New Zealand). Amplicon sequence variants (ASVs) were inferred using cutadapt v4.2 (Martin 2011) and DADA2 v1.32.0 (Callahan et al. 2016) within R (R Core Team 2024) using the parameters described in Pearman et al. (2022a). ASVs were taxonomically classified against the CyanoSeq (Lefler et al. 2023) version of the SILVA v138 database (Pruesse et al. 2007) using the RDP classifier (minBoot = 70; Wang et al. 2007) with eukaryotic, chloroplast and mitochondrial sequences removed from the dataset. Further processing of the data was undertaken in the phyloseq v1.48 package



(McMurdie and Holmes 2013). The number of reads present in the negative controls were removed from the samples via subtraction and ASVs that were present in only one replicate per lake were considered as potentially erroneous and removed from that lake's data. Metabarcoding data presented in this paper has been included previously in a national scale assessment (Pearman et al. 2022a, b).

For the metagenomic samples, DNA was extracted from 18 (six lakes in triplicate) surface sediment using an Alkaline Buffer Extraction Protocol (Thomson-Laing et al. 2022). Briefly, 3–5 g of sediment was transferred to 50 mL Falcon tubes (Corning Life Sciences) with 6 mL (0.33 M) sodium hydroxide and 3 mL (pH 6.7) Tris-EDTA buffer added to the samples. Samples were vortexed (20 sec) and incubated for 50 min at 65 °C. Subsequently, the samples were cooled for 5 min and centrifuged at 15 °C for one hour (15,000 × *g*). The resulting supernatant was transferred to a fresh tube and 7.5 mL of Tris HCl (1 M, pH 6.7), 1.5 mL sodium acetate (3 M, pH 5.2) and 100% molecular grade ethanol was added. To precipitate the DNA samples were stored at -20 °C overnight and centrifuged at 4 °C (10,000 × *g*) for one hour. The resulting pellet was transferred to PowerBead Tubes from a DNeasy PowerSoil Kit (Qiagen) and further DNA extraction was undertaken following the manufacturer's protocols using the automated Qiacube. Next generation library preparations were constructed following the VAHTS Universal Plus DNA Library Prep Kit instructions and were sequenced (2 × 150 bp) on an Illumina (San Diego, CA, USA) NovaSeq 6000 machine by Azenta Life Science (Suzhou, China).

Raw reads were trimmed using Trimmomatic (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:TRUE; LEADING:5; TRAILING:5; SLIDING WINDOW:10:15; MINLEN:30; Bolger et al. 2014). Individual samples were assembled using Megahit v1.2.9 (Li et al. 2015) using the meta-large preset. Contigs from each assembly were filtered using seqmagick v0.8.6 (Yu 2024) to greater than 500 base pairs (bp) and coding sequences were predicted using prodigal v2.6.3 (Hyatt et al. 2010). A non-redundant gene catalogue of Chatham Island coding sequences was constructed by concatenating the coding sequences from the individual assemblies and clustering them using CD-HIT-EST v4.8.1 (*c* = 0.95 -aS 0.9; Li and Godzik 2006) and further filtered to ensure coding sequences were greater than 100 bp. The quality cleaned raw reads were mapped against the non-redundant gene catalogue to obtain abundance scores using bowtie2 within sqm\_mapper.pl (Tamames and Puente-Sánchez 2019). The coding sequences were annotated using EggNOG-mapper (Huerta-Cepas et al. 2019; Cantalapiedra et al. 2021) using DIAMOND v2.1.8.162 (Buchfink et al. 2015). Taxonomic classification of the coding sequences was undertaken against NCBI's non-redundant (nr) database using blastp within DIAMOND (max target seqs = 10). The least common ancestor was determined for each coding sequence based on classifications that had bitscores within 10% of the highest bitscore for that sequence.

## Statistical analysis

Statistical analysis was undertaken in R (R Core Team 2024). For assessment of the taxonomic composition, the metabarcoding samples were merged at the lake level and rarefied to an even depth of 10,483 reads before transformation to relative abundance. The ASVs were grouped at the phyla level and visualization in ggplot (Wickham 2016) was undertaken for those phyla that accounted for on average greater than two percent of the community. Taxa that did not

reach this threshold or were not classified were amalgamated into “Other”. For the metagenomic dataset, the abundance of coding sequences was grouped at the phyla level and the data was merged to the lake level and relative abundance was calculated and plotted. To make the results comparable to the 16S rRNA gene data eukaryotes, viruses and unclassified sequences were removed for this analysis (these taxa were kept for all the other metagenomic analysis).

Multivariate analysis was undertaken on the unmerged metabarcoding dataset and the annotated metagenomic dataset. Bray-Curtis distance matrices were constructed and Principal Co-ordinate Analysis (PCoA) calculated within phyloseq and plotted in ggplot. Significant differences were assessed using PERMANOVA undertaken using the `adonis2` function in `vegan` v 2.6-8 (Oksanen et al. 2007).

Genes indicative of a variety of metabolic pathways were selected (Suppl. material 2) to investigate changes in metabolism among lakes. These selections were informed by literature searches and especially Yau et al. (2013). Linear regression was undertaken to assess relationships with environmental variables. For the linear regression sums of the genes for each selected pathway were found for each replicate and then the mean was found per lake. Differences in the abundance of osmolarity related genes was assessed with Kruskal-Wallis and pairwise comparisons assessed using the Dunn Test.

## Results

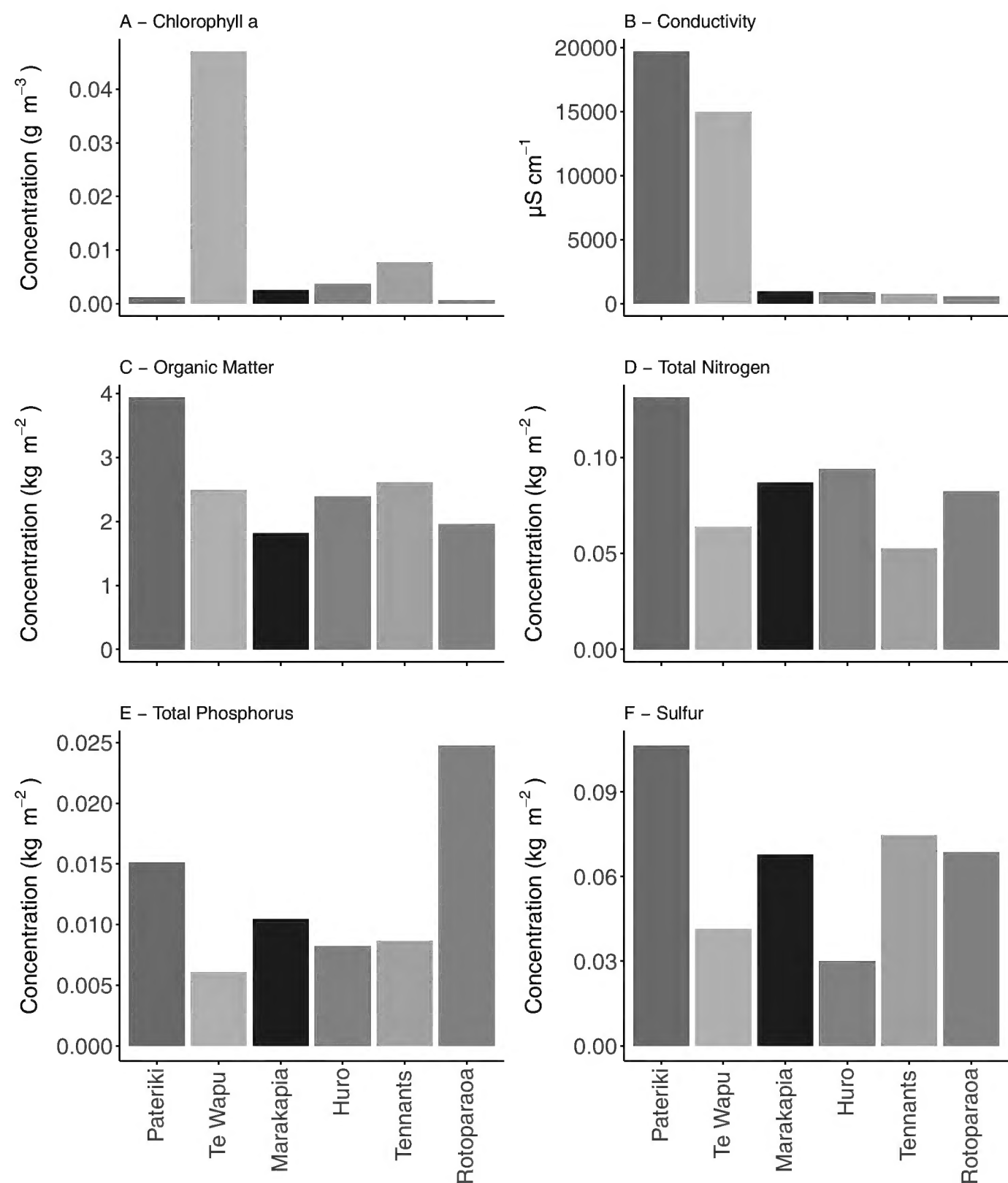
### Environmental data

Water chl-*a* levels were highest in Te Wapu reaching  $0.047 \text{ g m}^{-3}$  (Fig. 2A and Suppl. material 1). Two lakes had elevated conductivity levels (Pateriki  $19678 \mu\text{S cm}^{-1}$  and Te Wapu  $4978 \mu\text{S cm}^{-1}$ ; Fig. 2B and Suppl. material 1). Organic matter ( $3.94 \text{ kg m}^{-2}$ ), total nitrogen ( $0.131 \text{ kg m}^{-2}$ ) and sulfur ( $0.106 \text{ kg m}^{-2}$ ) were highest in the surface sediment of Pateriki whilst phosphorus was highest in Rotoparaoa (Fig. 2 and Suppl. material 1).

### Microbial composition and structure

Rarefaction analysis showed that sequencing depth was sufficient to sample the diversity of prokaryotic microbes in the lake sediments. However, after rarefaction (Suppl. material 5) at 10,483 reads (the minimum reads in a lake) a plateau was not reached in all samples but was sufficient to obtain an assessment of the main contributors to the community. The post rarefaction dataset contained a total of 3,859 ASVs with the alpha diversity results (Table 1) showing the highest richness being observed in Tennants (1565 ASVs) and lowest in Pateriki (300 ASVs). Only one ASV belonging to Nitrospirota was found in all lakes. The metagenomic dataset resulted in a total of 12,111,534 contigs after filtering (Suppl. material 3 for details). After processing, the Chatham Island gene catalogue resulted in 11,785,220 coding sequences. Of these 6,868,877 could be annotated by EggNOG-mapper and 4,097,053 had KEGG annotations.

Multivariate analysis showed that there was a significant difference ( $F = 17.04$ ;  $p < 0.001$ ) in the sediment community composition among the lakes. The two lakes with saline influences (Te Wapu and Pateriki) were the most distinct

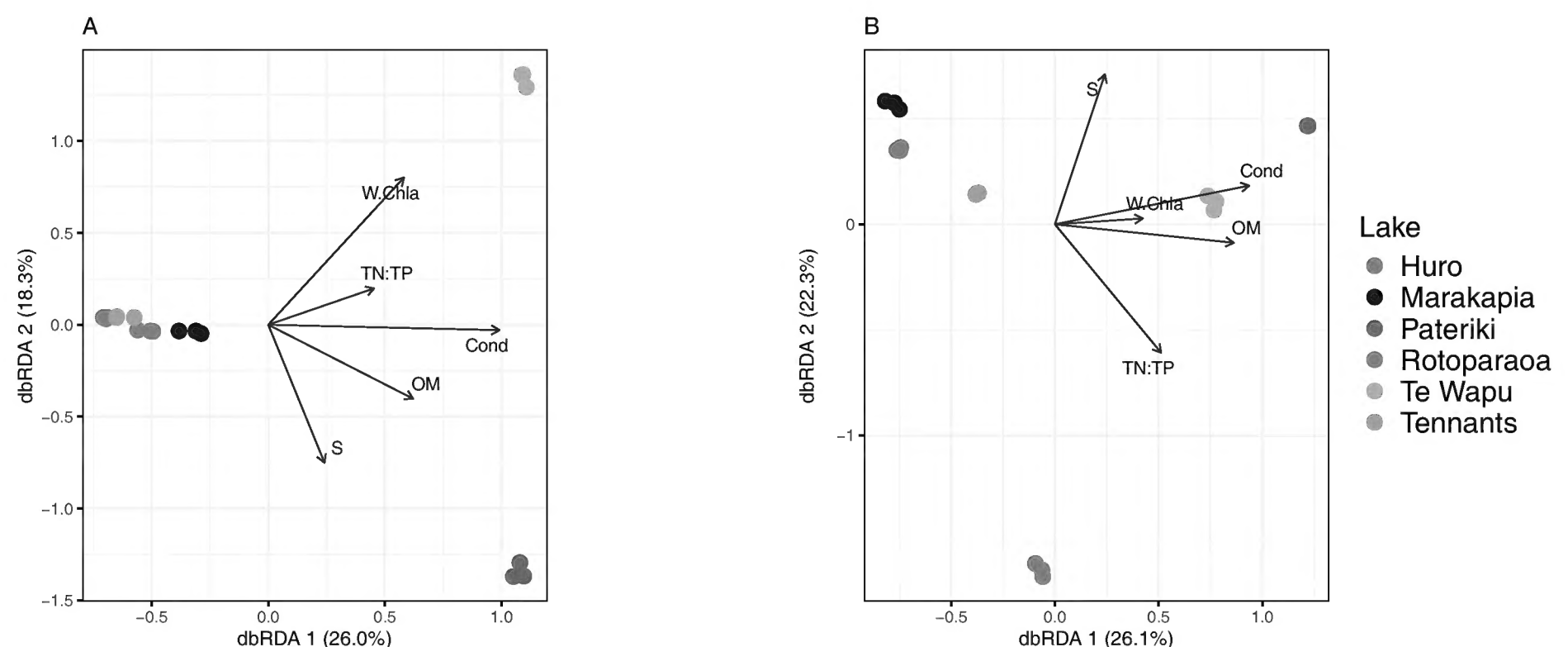


**Figure 2.** Environmental data the study lakes on Chatham Island. Water column. **A** Chlorophyll *a* and **B** conductivity. Surface sediment. **C** Organic material. **D** Total nitrogen. **E** Phosphorus. **F** Sulfur.

**Table 1.** Alpha diversity metrics for the six lakes on Chatham Island. ASV = amplicon sequence variant.

Lake	Observed ASVs	Shannon Diversity
Huro	378	4.5
Marakapia	370	5.3
Pateriki	300	5.2
Rotoparaoa	970	5.9
Te Wapu	1275	6.2
Tennants	1565	6.8

(Fig. 3A). The functional analysis showed similar significant differences amongst lakes ( $F = 21.842$ ;  $p < 0.001$ ; Fig. 3B) with mantel tests showing a strong and significant correlation (Spearman’s rank correlation  $\rho = 0.586$ ;  $p = 0.001$ ) between the taxonomic and functional distance matrices. Of note in the metagenomic analysis was the distinct difference of Huro from all other lakes (Fig. 3B).



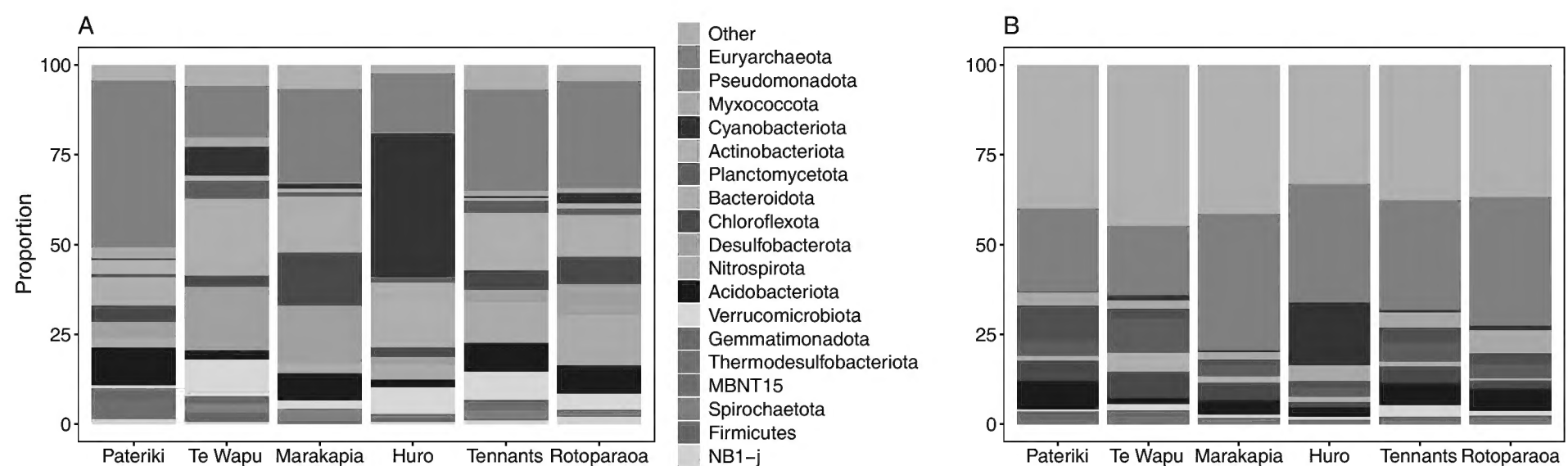
**Figure 3.** Distance based redundancy analysis for the study lakes on Chatham Island for **A** Community composition results from metabarcoding and **B** Functional composition from metagenomics. TN:TP = Ratio of total nitrogen and total phosphorus; OM = Organic matter; W.Chla = Chlorophyll a in the water column; S = Sulfur; Cond = Conductivity.

Metabarcoding showed that, overall, Pseudomonadota were the predominant phyla (14.5% in Te Wapu to 46.5% in Pateriki; Fig. 4A; Suppl. material 4 for ASV table). This was mostly comprised of Gammaproteobacteria (mean = 22.4%) although Alphaproteobacteria (14.8%) contributed to the community in Pateriki. In Huro, Cyanobacteriota (genus *Cyanobium*) contributed 40.3% of the community with Bacteroidota also highly abundant in the metabarcoding results (8.1% in Pateriki to 21.6% in Te Wapu). In the metagenomics bacteria accounted for 90% with 7% not being able to be classified (contig table [<https://doi.org/10.6084/m9.figshare.28465397>]). The remainder of the composition contained archaea (2%), eukaryotes (0.4%) and viruses (0.1%). In a comparison with the metabarcoding data the prokaryotes in the metagenomics (Fig. 4B) were assessed, with Pseudomonadota as the predominant classified phyla (19.2% in Te Wapu to 35.8% in Rotoparaoa). This was mainly comprised of Gammaproteobacteria (2.3% in Huro to 16.4% in Marakapia), Alphaproteobacteria (1.0% in Marakapia to 10.8% in Pateriki) and Betaproteobacteria (2.8% in Pateriki to 15.3% in Huro). Cyanobacteria were important components in Huro in agreement with the metabarcoding results accounting for 17.6% of the coding sequences. Myxococcota (mean 3.9%), Planctomycetota (mean 5.0%), Acidobacteriota (mean 4.9%) and Actinomycetota (mean 3.7%) comprised small proportions of the community across all lakes based on the metagenomics.

## Metabolic functioning

Differences were observed in metabolic functions among lakes (Fig. 5). Various pathways for nitrogen metabolism were investigated and correlated with nitrogen related environmental variables. The highest values for nitrogen fixation, dissimilatory nitrate reduction to ammonia (DNRA) and nitrification were in Rotoparaoa while Pateriki had the highest values for assimilatory nitrate reduction and denitrification (Fig. 5). In general, there was a negative trend between nitrogen metabolism and surface sediment TN:TP ratio although this was only significant for DNRA ( $r^2 = 0.779$ ;  $p = 0.01$ ; Fig. 6B) and nitrification ( $r^2 = 0.928$ ;  $p = 0.002$ ; Fig. 6E). Across





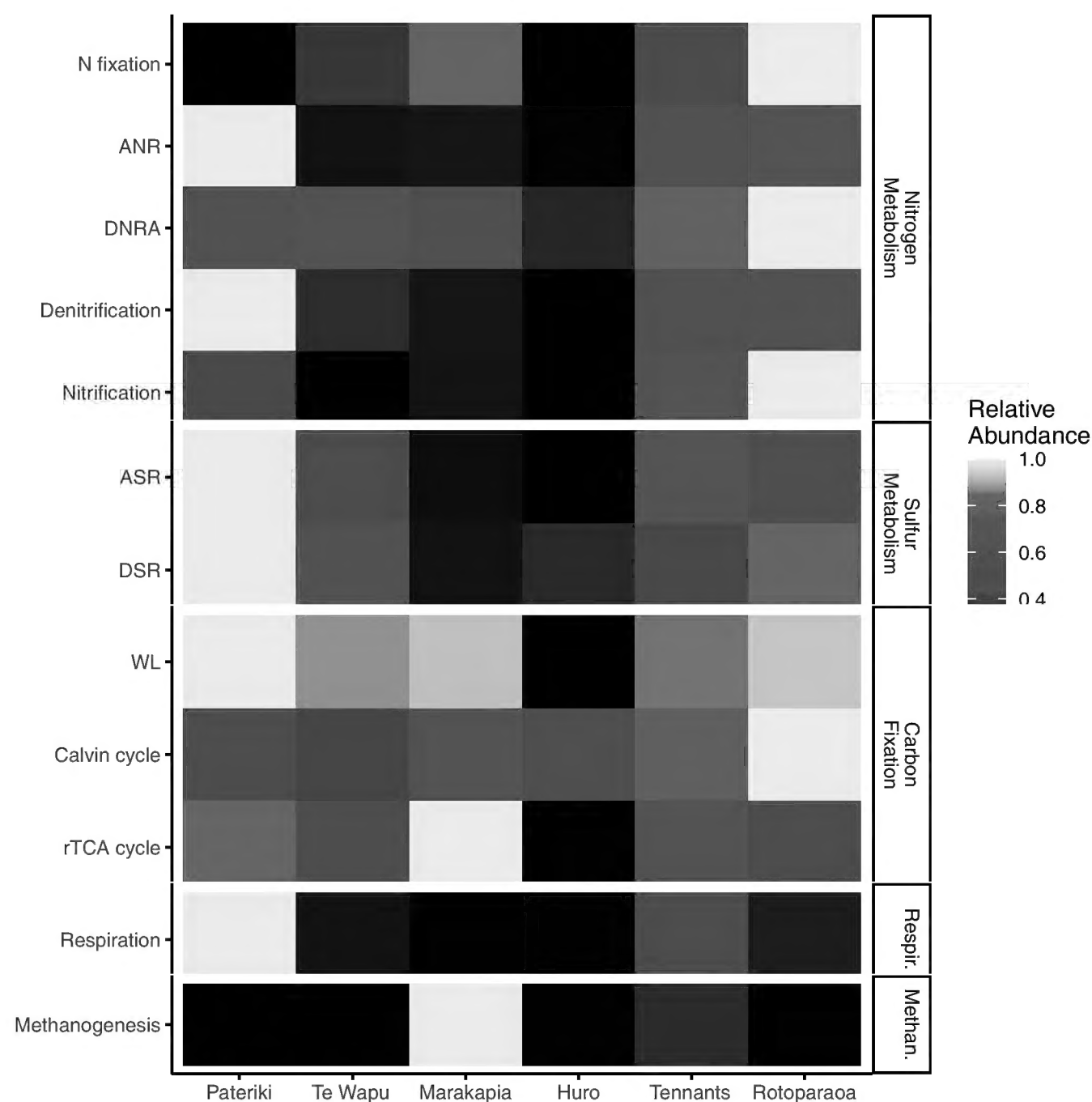
**Figure 4.** Taxonomic composition at the phylum level in the study lakes on Chatham Island for **A** 16S rRNA gene metabarcoding, and **B** Metagenomic coding sequences. 'Other' consists of taxa not assigned at the Phylum level as well as those that accounted for less than 2% of the community in at least one lake.

the nitrogen cycling pathways Betaproteobacteria and Gammaproteobacteria generally had the highest contribution of those counts that could be assigned at the class level. For nitrification however, Nitrospira was the most predominant taxa across the lakes (<https://doi.org/10.6084/m9.figshare.28465397>). In Huro, the methanogenic archaea Methanomicrobia accounted for the highest contribution to nitrogen fixation (<https://doi.org/10.6084/m9.figshare.28465397>).

For sulfur metabolism, the highest abundances for assimilatory sulfate reduction were observed in Pateriki and there was a strong and significant positive relationship to the amount of sulfate in the sediment ( $r^2 = 0.744$ ;  $p = 0.027$ ; Fig. 7A). When compared against the amount of organic matter there was a weaker positive relationship that was not significant ( $r^2 = 0.593$ ;  $p = 0.073$ ; Fig. 7B). For dissimilatory sulfate reduction positive relationships were observed although these were weaker and not significant for both sediment sulfate ( $r^2 = 0.372$ ;  $p = 0.214$ ; Fig. 7C) and organic matter ( $r^2 = 0.400$ ;  $p = 0.18$ ; Fig. 7D). Taxonomic classification of both the sulfate reduction pathways showed that in Pateriki, Alphaproteobacteria (phylum Pseudomonadota) had the highest contribution while Betaproteobacteria and Gammaproteobacteria were more prevalent in the other lakes with Cyanophyceae (Cyanobacteriota) contributing in Huro (<https://doi.org/10.6084/m9.figshare.28465397>).

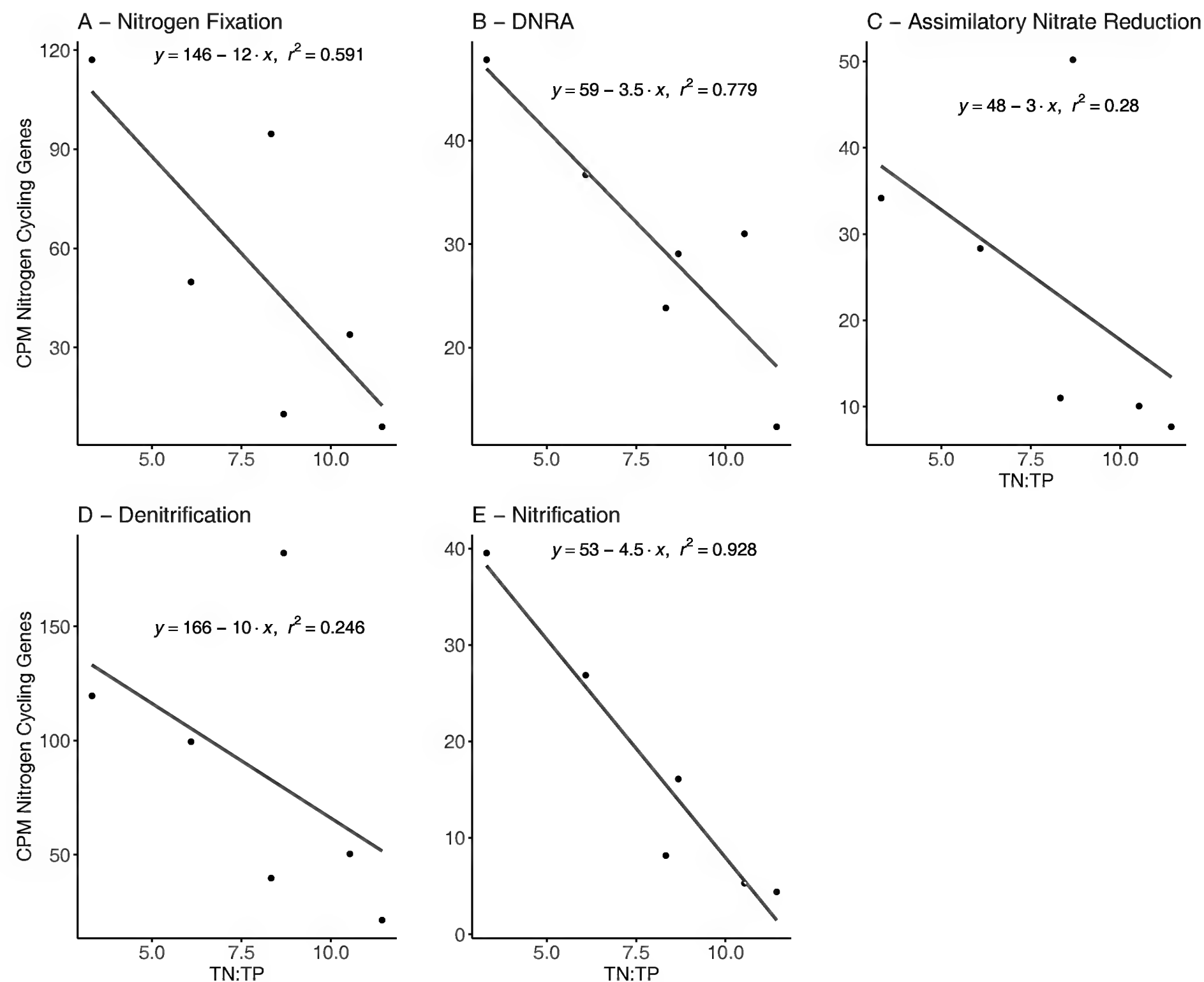
In four of the six lakes the prevalence of methogenesis indicator genes (K00400 + K00401) was negligible with less than five copies per million. The highest value was observed in Marakapia with 88 copies per million (Fig. 5).

Genes indicative of the Calvin cycle were prevalent across the lakes and predominantly belonged to Pseudomonadota, especially Betaproteobacteria, although in Marakapia there was a contribution from Euryarchaeota while in Huro, Cyanobacteriota accounted for an increased proportion compared to the other lakes (Fig. 8A). Genes for the reductive TCA cycle were also observed in the lakes and of those that could be classified the phylum Nitrospirota was observed to be a major component in Pateriki, Tennants and Rotoparaoa while Archaea accounted for a higher proportion in Marakapia. Thermodesulfobacteriota contributed to the rTCA community in Pateriki and Te Wapu (Fig. 8B). Investigation of the Wood-Ljungdahl pathway using acetyl-CoA synthase alpha (K14138) subunit as an indicator showed high abundance in Pateriki and low in Huro. While the majority of the coding sequences were unclassified, Chloroflexota had the highest number of classified acetyl-CoA synthase genes (Fig. 8C).



**Figure 5.** Distribution of functions in the in the study lakes on Chatham Island. The relative abundance of each function is scaled to the highest value. KEGG IDs used for the characterization of functions are detailed in Suppl. material 2. rTCA = reductive Tri-carboxylic Acid cycle, WL = Wood-Ljungdahl pathway, DSR = Dissimilatory Sulfate Reduction, ASR = Assimilatory Sulfate Reduction, ANR = Assimilatory Nitrate Reduction, DNRA = Dissimilatory Nitrate Reduction to Ammonia, N = nitrogen.

The production of osmoprotectants (glycine betaine [betA, betB and betI] and proline [proA, proB and proC]) were significantly different amongst lakes and in general higher in Pateriki than other lakes. However, pairwise testing (Dunn Test) revealed Pateriki only had significant differences ( $p < 0.05$ ) with Huro and Marakapia for glycine betaine (Fig. 9A; Suppl. material 2) and between Pateriki and Huro for proline (Fig. 9B; Suppl. material 2). In Pateriki and Te Wapu the glycine betaine genes were predominantly classified as Pseudomonadota with the class Alphaproteobacteria particularly prominent while in the lakes with no saline influence Betaproteobacteria was more prevalent (<https://doi.org/10.6084/m9.figshare.28465397>). Alphaproteobacteria were also prominent contributors for proline in Pateriki but the other lakes had a mix of Beta- and Gamma- proteobacteria with cyanobacteria also contributing in Huro. Similarly, the genes encoding transporters linked to osmoprotection (proP, proV, proW, proX, betT, opuABCDE) were more abundant in Pateriki (Fig. 9C; Suppl. material 2) compared to the other lakes although pairwise tests indicated only the Pateriki – Huro comparison was significantly different ( $p < 0.05$ ). Taxonomic classifications at the class level for osmotransporters were predominantly related to Alphaproteobacteria across the lakes except for Marakapia and Rotoparaoa where Gammaproteobacteria and



**Figure 6.** Regression analysis of indicative genes against the Total Nitrogen:Total Phosphorus (TN:TP) ratio for **A** Nitrogen fixation. **B** Dissimilatory nitrate reduction to ammonia. **C** Assimilatory nitrate reduction. **D** Denitrification. **E** Nitrification. See Suppl. material 2 for genes used. CPM = counts per million. Note y-axis scales differ. Note CPM across plots should not be compared as a different number of indicative genes were used for each pathway. Gene abundances were summed for each pathway.

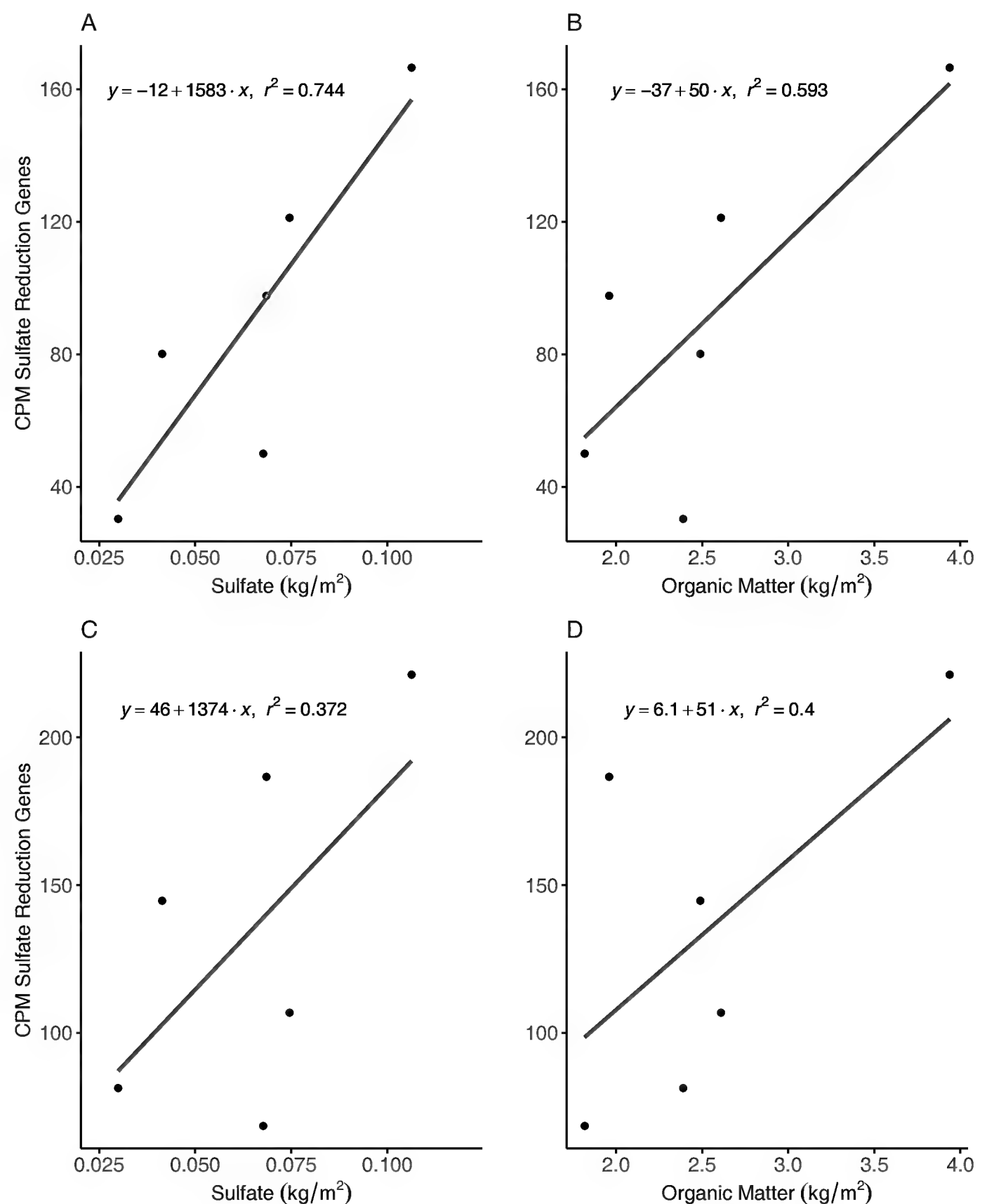
Vicinamibacteria (phylum Acidobacteriota) had higher count numbers (<https://doi.org/10.6084/m9.figshare.28465397>). Similar patterns for increased abundances in Pateriki for potassium transporters and trehalose synthesis were also observed.

Discussion

Coastal dune lakes are globally rare, and little is known about the metabolic processes that occur within them. By assessing six lakes on Chatham Island where there are no non-native fish or macrophytes, and relatively low levels of human impact, a better understanding of differences in the metabolic functioning of these lakes has been garnered.

Composition

In both the metabarcoding and the metagenomics data Pseudomonadota were dominant, a finding which has also been shown in previous lake sediment

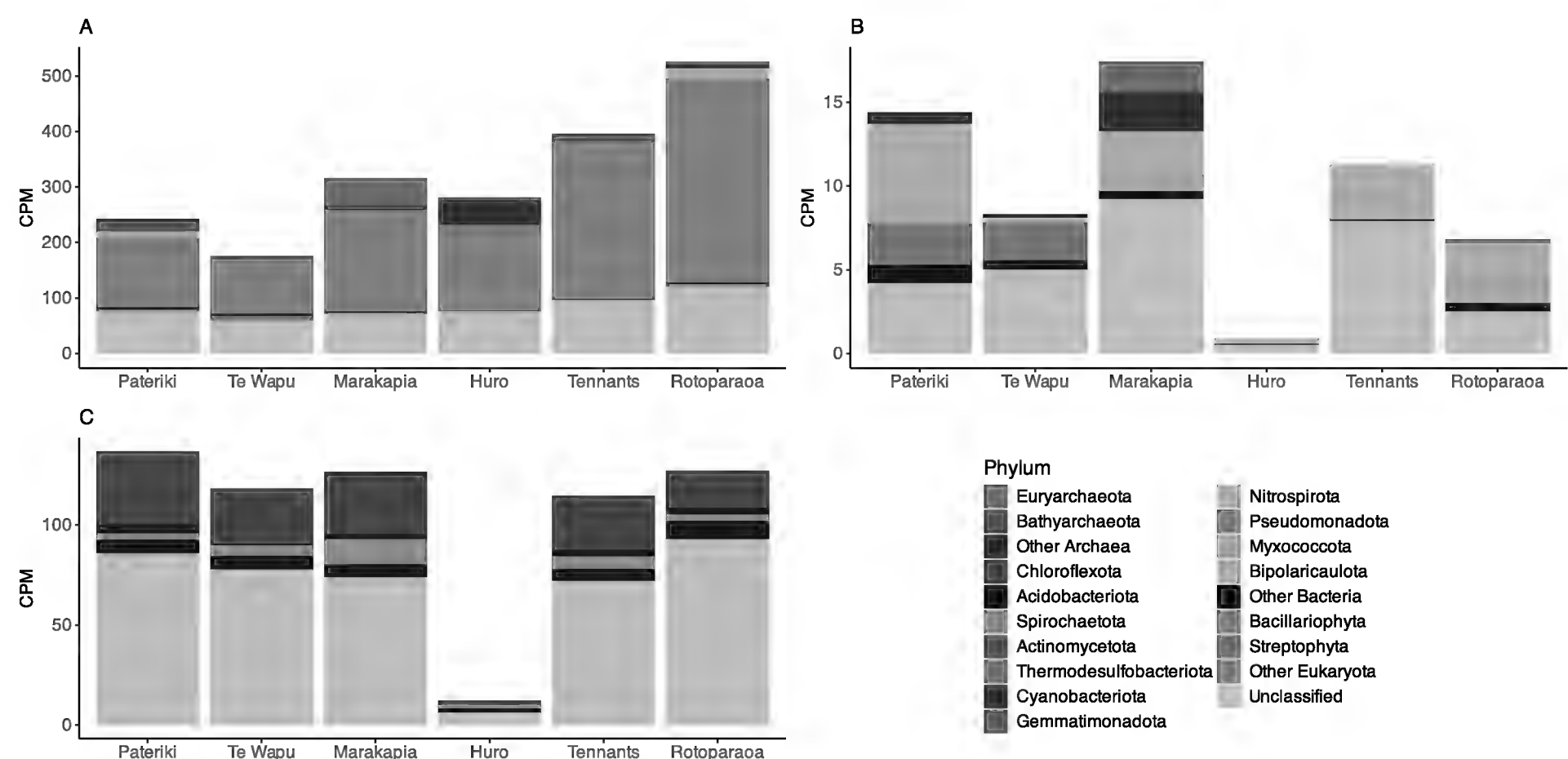


**Figure 7.** Regression analysis of indicative genes for assimilatory sulfate reduction against surface sediment sulfate concentrations (A), and organic matter (B). Dissimilatory sulfate reduction against surface sediment sulfate concentrations (C), and organic matter (D). CPM = counts per million. Note y-axis scales differ. Note CPM across plots should not be compared as a different number of indicative genes was used for each pathway.

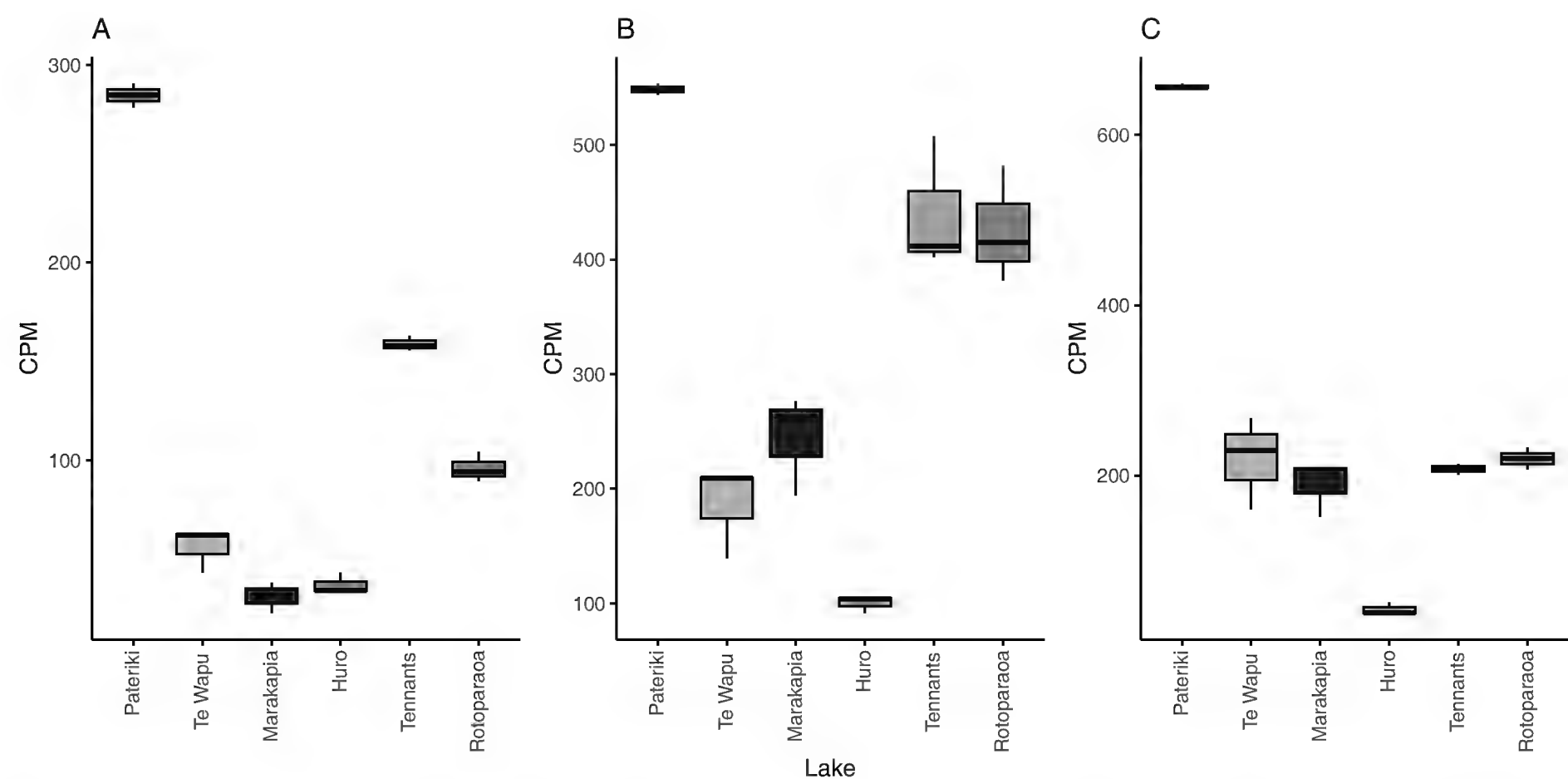
studies (Zhang et al. 2015; Pearman et al. 2020; Biessy et al. 2022; Karlicki et al. 2024). Pseudomonadota are actively involved in a variety of biogeochemical pathways in aquatic ecosystems (Song et al. 2012) including nitrogen fixation, photosynthesis, sulfur and methane metabolism and the removal of nitrogen and phosphorus from the sediment (Kuang et al. 2023). Within the Pseudomonadota, Alphaproteobacteria were more predominant in the saline influenced lakes with Betaproteobacteria accounting for higher proportions in the freshwater lakes. This is in agreement with previous findings that show that salinity can influence the composition with Betaproteobacteria being intolerant of haline environments (Lew et al. 2022).

In Huro, a large proportion of the taxonomic classifications are assigned to Cyanobacteria, with a smaller proportion present in Te Wapu and negligible amounts





**Figure 8.** Counts per million (CPM) of indicative genes for the Calvin cycle (A), reductive Tricarboxylic Acid cycle (B) and Wood-Ljungdahl (C) coloured by taxonomic classifications. See Suppl. material 2 for indicative genes.



**Figure 9.** Counts per million (CPM) of genes related to osmoprotection. **A** Glycine Betaine metabolism genes. **B** Proline metabolism genes. **C** Osmoprotection transporters.

in the other lakes. The metabarcoding results showed that the Cyanobacteria community in Huro comprised the picocyanobacterial genus *Cyanobium* which has previously been observed as a major contributor in freshwater lakes (Li et al. 2020; Schallenberg et al. 2021b). Picoeukaryotes have previously been shown to be a major component of lakes (Schallenberg et al. 2021a) although there is often a negative correlation between picocyanobacterial abundance and trophic state as picoeukaryotes become more competitive (Ruber et al. 2018). However, large abundances of picocyanobacteria have been noted in more eutrophic lakes especially during summer blooms (Schallenberg et al. 2021b).

There were differences in the phyla that were classified based on metabarcoding and metagenomics, a finding also observed in a study on the Ōtuwharekai (Ashburton) lakes in Aotearoa New Zealand (Biessy et al. 2022). A key difference was the prevalence of Nitrospirota and Desulfobacterota in the metabarcoding results. There could be several reasons for this discrepancy. Firstly, in this study the metabarcoding protocol targets the 16S rRNA gene which is a multiple-copy gene and this may bias the results towards taxa with higher copy numbers (Kembel et al. 2012). Secondly, there is a limited number of annotated genomes currently present in reference databases which affects the ability of genes from metagenomes to be correctly identified. This is also highlighted by the high proportion of coding sequences that could not be annotated by the least common ancestor approach to a phylum. Different approaches to taxonomic classification such as assigning taxonomy to the raw reads or contigs, especially in the context of metagenomic assembled genomes, could improve the classification results and enable more in-depth assessment of what organisms are undertaking the functions within the environment. With the increasing use of genomics in environmental studies and more genomes being incorporated into reference databases, there is likely to be an increased ability to phylogenetically classify metagenomic reads in the future.

Multivariate analysis of the community lakes showed the two lakes with saline influences were distinct from the others. A similar pattern was also observed for the functioning of the microbial community in the lakes. Salinity has been shown previously to be an important environmental driver of variation in microbial communities (Cabello-Yeves et al. 2022; Banchi et al. 2024; Wu et al. 2024). For example, as noted in this study there is a decrease in Betaproteobacteria in the saline- influenced lakes and it is likely that salinity is driving at least in part the changes in composition and function of the lake communities. However, the composition of microbial communities in lakes is a combination of local and regional scale factors (Lindström and Langenheder 2012) and it should be noted that these two lakes are also separated from the other lakes on the north-east of the island and differences due to geographic distance and factors such as dispersal limitation cannot be ruled out.

## Functional assessment

### Nitrogen metabolism

The lakes on Chatham Island are naturally enriched with phosphate (Meredith and Croucher 2007) and in general, have limited inputs of nitrogen potentially leading lakes to be nitrogen-limited. Across the lakes there was a negative trend between the abundance of nitrogen cycling-related genes (e.g., DNRA and nitrification) and the TN:TP ratio. This indicates that the abundance of these genes, and thus functional potential, is higher when there is greater nitrogen limitation in the lake sediment. Previous work in terrestrial environments has shown that DNRA limits the loss of nitrate from the system and outcompetes denitrification in nitrate-limited conditions (Pandey et al. 2020) while denitrification rates have been shown to rapidly increase in eutrophic systems (Jiang et al. 2023). This would indicate that there is a higher potential of DNRA and keeping nitrate within the system when nitrogen is limited while this pathway in the metagenome has lower copies per million when there is no limitation, and the conservation of nitrogen is less

important. However, metagenomics only shows the functional potential in the community and further work to measure rates and gene expression is required to fully understand nitrogen cycling within the dune lakes of Rēkohu.

### Sulfur metabolism

Sulfur cycling plays a key role in aquatic ecosystems and is vital in many interconnected pathways. In general, sulfate levels in seawater are higher compared to those in freshwater (Holmer and Storkholm 2001; Margalef-Marti et al. 2023) and in the current study, Pateriki which is impacted by saline influences has higher concentrations of sulfur in the sediment. The availability of sulfate in the environment can affect the remineralization of organic matter in sediments with sulfate reduction, often outcompeting methanogenesis in sediments with elevated concentrations of sulfur (Holmer and Storkholm 2001; Margalef-Marti et al. 2023). In the current study, there were significant positive relationships between assimilatory sulfate reduction and the concentration of sulfur in the sediment while dissimilatory sulfate reduction had a similar, albeit non-significant pattern. This is in agreement with other work in lakes of Aotearoa-New Zealand (Biessy et al. 2022). Sulfur-bearing organic matter from both allochthonous terrestrial and autochthonous primary production is often the dominant component of the total sulfur pool (Holmer and Storkholm 2001; Couture et al. 2016) and is likely to be a substantial component in the lakes of the Chatham Islands. In general, the metagenomics indicated that methanogenesis was most prevalent in Marakapia where sulfate reduction was low suggesting that in this lake methanogenic organisms contribute more substantially to organic remineralization compared to the other lakes. The increased presence of coding sequences that belong to Euryarchaeota in the metagenomics of Marakapia, which include methogenic members, would reinforce the possibility of higher amounts of methanogenesis in this lake.

### Carbon fixation

Different carbon fixation strategies are prevalent amongst microbes and the analysis of functional genes indicative of key steps provides a way to investigate the strategies of inorganic carbon assimilation in environmental samples (Hügler and Sievert 2011). The enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) is a key enzyme in the Calvin-Benson-Bassham (CBB) cycle and is used as a marker for autotrophic organisms. In the current study, most of the reads mapped to RubisCO genes were related to Pseudomonadota, with the majority belonging to Betaproteobacteria with fewer contributions from Gammaproteobacteria and Alphaproteobacteria. This finding is similar to that of Alfreider et al. (2017) who found chemoautotrophic Betaproteobacteria were prevalent in stratified lakes in Europe. While most of the RubisCO reads in the current study belonged to chemoautotrophs, a smaller number were related to photoautotrophs especially Cyanobacteria in Huro as well as small numbers of the eukaryotes Bacillariophyta and Streptophyta in Te Wapu. This indicates that despite the lakes being shallow (four lakes < 1.5 m) photoautotrophs are not a major component of the benthic community and that there is limited deposition of phytoplankton from the water column. This may be linked to lower light availability due to the tea stained color of the waters of the lakes.

Genes for the reductive tricarboxylic acid cycle (rTCA) cycle were present across the lakes, although minimal in Huro, but this pathway likely constitutes only a minor part of carbon fixation within the lakes. Previous evidence has shown that this pathway has been observed in a variety of taxa across the bacterial and archaeal tree of life (Garritano et al. 2022). In the present study, the classification of the rTCA genes varied across lakes with Nitrospirota the major contributor in Pateriki, Tennants and Rotoparaoa. Indicative genes for the rTCA cycle have previously been observed in lakes, albeit mainly in anoxic regions of the water column (Alfreider et al. 2017). The rTCA pathway has predominantly been found in anaerobic organisms although enzyme modifications to allow functionality in oxic environments are known (Yamamoto et al. 2006) including in members of the Nitrospirota (Lücker et al. 2010). Thermodesulfobacteriota were prevalent in the lakes with saline influences (Pateriki and Te Wapu) which have previously been shown to contribute to carbon fixation via rTCA in both lakes and other environments (Fang et al. 2022; Angius et al. 2024) and are likely relying on the oxidation of sulfur compounds for energy generation (Garritano et al. 2022). Recent work supports the idea that species within the archaea phylum Thermoplasmatota can undertake carbon fixation via the rTCA pathway (Garritano et al. 2022). The classification of key genes relating to the rTCA pathway in Marakapia highlights the potential of archaea to use this pathway for carbon fixation. Further work, including deeper sequencing and the construction of metagenomic assembled genomes (MAGs) from this lake, could provide more evidence for the rTCA pathway being present in archaea.

The Wood-Ljungdahl pathway enables acetogens to convert  $H_2$  and  $CO_2$  anaerobically into acetyl-CoA (Hügler and Sievert 2011). In the current study, most sequences relating to this pathway were unable to be classified, however of those that were taxonomically assigned a high proportion belonged to Chloroflexota. Previous work has shown members of the class Dehalococcoidales contain the Wood-Ljungdahl pathway in anoxic zones of aquatic environments (Chen et al. 2023). Enzymes involved in the Wood-Ljungdahl pathway are negatively affected by oxygen, suggesting that these pathways would be restricted to anoxic sediment. Lake sediment is often characterized by a steep oxygen gradient, and the organisms undertaking the Wood-Ljungdahl pathway would likely be found deeper in the sediment, however, the amalgamation of the top two centimeters of sediment for samples in the current study negates the identification of depth gradients and further work would be required. Previous work has shown that salinity has an effect on energy metabolism with the Wood-Ljungdahl pathway possibly providing an advantage in higher salinities (Fang et al. 2022). This could explain the trends observed in the Rēkohu lakes where the abundance of the indicative genes for the CBB cycle in Pateriki, where salinity was highest, was lower than in the other lakes whilst those for the Wood-Ljungdahl pathway were higher in Pateriki.

## Osmoprotection

Increases in salinity in the environment can cause osmotic stress and lead to the loss of cell turgor without a cellular response (Gouffi and Blanco 2000). One of the main responses to increased osmotic stress is the production or uptake of bacterial-compatible solutes (such as glycine betaine and proline). In the current



study, two lakes (Te Wapu and especially Pateriki) had a saline influence, and we analyzed the metagenomes to assess if differences were evident in gene abundances related to synthesis of the compatible solutes' glycine betaine and proline as well as the presence of compatible solute transporters (Fig. 9). Genes for the synthesis of glycine betaine were more abundant in Pateriki than the other lakes with levels twice those found in the other lakes although pairwise comparisons in general showed no significant difference between the lakes. The lack of significant difference could be due to a lack of statistical power and further work with more replicates would be required to improve the power of the statistical tests. However, the increased abundance of these genes in Pateriki would indicate that microbes are utilizing glycine betaine as an osmoprotectant although no evidence was seen of an elevated response in Te Wapu. There were also shifts in the taxonomic classifications related to the production of glycine betaine with both Pateriki and Te Wapu both having higher contributions from Alphaproteobacteria while other lakes had a higher contribution for Betaproteobacteria. This would be in agreement with findings that show that Betaproteobacteria are intolerant of haline stress and are thus restricted in aquatic environments to freshwater habitats (Lew et al. 2022) and similar trends were observed in this study for the other osmolarity related genes. The trends for proline were not as clear as for the metabolism of glycine betaine as counts per million in Pateriki were not much higher than those in Tennants and Rotoparaoa which had no saline influence. However, proline is known to be produced in response to multiple other environmental stresses (e.g. pH stress, metal toxicity, ultraviolet radiation; Goswami et al. 2022) as well as in carbon and nitrogen metabolism (Goswami et al. 2022). In the current study, proline is likely being used for other metabolic functions in Tennants and Rotoparaoa at least and is thus not an indicator of functional potential adaptation to higher osmolarity in these lakes. For the compatible solute transporters, higher abundances were observed for Pateriki compared to the other lakes. Despite Te Wapu having elevated conductivity, unlike Pateriki no increased abundance of genes in response to this osmotic stress was observed in the sediment. While metagenomic results only give an indication of the functional potential of the community and not the actual production of the proteins, changes could occur transcriptionally in Te Wapu. However, as there was a substantial response for Pateriki there could be other reasons for this lack of response in Te Wapu. The bacteria may be using other mechanisms to mitigate the effects of salinity. However, no trend was observed for other mechanisms of osmoprotection such as for trehalose synthesis or potassium transporters (Wood 2015) in Te Wapu (but both had elevated abundances in Pateriki; data not shown). Another possible reason could be that there is variation in the salinity through the water column in Te Wapu and the sediment does not experience the same salinity conditions as measured at the surface. Monitoring data in Te Wapu has shown that salinity is temporally variable (Environment Canterbury 2024) and further investigations of spatial and temporal variability in salinity are required.

## Conclusions

This investigation provided an increased understanding of the metabolic functioning of microbial communities in globally rare coastal dune lakes. Multivariate analysis showed the effect of salinity on both the composition and

functioning of the lakes. The results showed that there were distinct patterns in nitrogen and sulfur cycling genes against nutrient conditions within the lake. The results also highlighted the strategies of microbes to osmolarity stress with increased abundance of osmoprotectants especially in the alphaproteobacteria. With lakes coming under increasing pressures, having a deeper understanding of how lakes function will enable more informed decision-making processes in the preservation of lake ecosystems. With increased understanding of how key marker genes respond both temporally and spatially could provide vital information to assess changes in the health of a lake from various pressures (e.g. eutrophication, antibiotic-resistant genes, heavy metals (Nandy and Kapley 2024) and indicate overall trends (Rohwer et al. 2025) which can inform management decisions. These data provide valuable information that can also provide insights into ecosystem wide changes, for example, the rationale for why cyanobacterial or other algal blooms occur in some, but not all, of these lakes during summer, as well as differences in water quality.

## Acknowledgements

We thank all members of the Lakes380 team ([www.ourlakesourfuture.co.nz/lakes/](http://www.ourlakesourfuture.co.nz/lakes/)) for field assistance, Sean Waters for the nutrient data and Lena Schallenberg (Cawthron) for producing Fig. 1. We thank Hokotehi Moriori Trust, Ngāti Mutunga o Wharekauri Trust and Environment Canterbury for their support of this project.

## Additional information

### Conflict of interest

The authors have declared that no competing interests exist.

### Ethical statement

No ethical statement was reported.

### Funding

This research was funded by the New Zealand Ministry of Business, Innovation and Employment Smart Idea project – “Applying a functional evidence approach to prioritise lake restoration initiatives (CAWX2303)” and Research programs – “Our Lakes, Our Future (CAWX2305)” and Our lakes’ health; past, present, future (C05X1707).

### Author contributions

Conceptualization: SAW, JKP. Data curation: JKP. Formal analysis: JKP. Funding acquisition: JKP. Methodology: GTL, JKP, JS, JKK. Writing – original draft: SAW, JKP. Writing – review and editing: JS, GTL, JKP, SAW, JKK.

### Author ORCIDs

John K. Pearman  <https://orcid.org/0000-0002-2237-9723>

Joseph Kanyi Kihika  <https://orcid.org/0000-0001-6805-5890>

Georgia Thomson-Laing  <https://orcid.org/0000-0001-8337-5489>

Susanna A. Wood  <https://orcid.org/0000-0003-1976-8266>

## Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

Raw metabarcoding sequences for the lakes in the current study are stored as part of a larger investigation in the NCBI sequence read archive (SRA) under accession PRJNA813318. Raw files for the metagenomic data used in this study are deposited in the Genomics Aotearoa Data Repository at: <https://doi.org/10.57748/3jdy-rc84> and are a subset of a larger study.

## References

- Alfreider A, Baumer A, Bogensperger T, Posch T, Salcher MM, Summerer M (2017) CO<sub>2</sub> assimilation strategies in stratified lakes: Diversity and distribution patterns of chemolithoautotrophs. *Environmental Microbiology* 19: 2754–2768. <https://doi.org/10.1111/1462-2920.13786>
- Angius F, Cremers G, Frank J, Witkowski C, Pol A, van Alen TA, Jetten MSM, Op den Camp HJM, Berben T (2024) Gene-centered metagenome analysis of Vulcano Island soil (Aeolian archipelago, Italy) reveals diverse microbial key players in methane, hydrogen and sulfur cycles. *Antonie van Leeuwenhoek* 117: 94. <https://doi.org/10.1007/s10482-024-01995-5>
- Banchi E, Corre E, Del Negro P, Celussi M, Malfatti F (2024) Genome-resolved metagenomics of Venice Lagoon surface sediment bacteria reveals high biosynthetic potential and metabolic plasticity as successful strategies in an impacted environment. *Marine Life Science & Technology* 6: 126–142. <https://doi.org/10.1007/s42995-023-00192-z>
- Biessy L, Pearman JK, Waters S, Vandergoes MJ, Wood SA (2022) Metagenomic insights to the functional potential of sediment microbial communities in freshwater lakes. *Metabarcoding and Metagenomics* 6: e79265. <https://doi.org/10.3897/mbmg.6.79265>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (Oxford, England) 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60. <https://doi.org/10.1038/nmeth.3176>
- Cabello-Yeves PJ, Callieri C, Picazo A, Schallenberg L, Huber P, Roda-Garcia JJ, Bartosiewicz M, Belykh OI, Tikhonova IV, Torcello-Requena A, De Prado PM, Puxty RJ, Millard AD, Camacho A, Rodriguez-Valera F, Scanlan DJ (2022) Elucidating the picocyanobacteria salinity divide through ecogenomics of new freshwater isolates. *BMC Biology* 20: 175. <https://doi.org/10.1186/s12915-022-01379-z>
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13: 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38: 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Champion P, Clayton JS (2004) Aquatic vegetation of Chatham Island (Rekohu). Department of Conservation, Wellington.

- Chen X, Liu J, Zhu X-Y, Xue C-X, Yao P, Fu L, Yang Z, Sun K, Yu M, Wang X, Zhang X-H (2023) Phylogenetically and metabolically diverse autotrophs in the world's deepest blue hole. *ISME Communications* 3: 117. <https://doi.org/10.1038/s43705-023-00327-4>
- Couture R-M, Fischer R, Van Cappellen P, Gobeil C (2016) Non-steady state diagenesis of organic and inorganic sulfur in lake sediments. *Geochimica et Cosmochimica Acta* 194: 15–33. <https://doi.org/10.1016/j.gca.2016.08.029>
- Downing JA, McCauley E (1992) The nitrogen: Phosphorus relationship in lakes. *Limnology and Oceanography* 37: 936–945. <https://doi.org/10.4319/lo.1992.37.5.0936>
- Environment Canterbury (2024) ECan Water quality data. Environment Canterbury. <https://www.ecan.govt.nz/data/water-quality-data/wqdetails/?SiteID=SQ34887> [September 20, 2024]
- Fang Y, Liu J, Yang J, Wu G, Hua Z, Dong H, Hedlund BP, Baker BJ, Jiang H (2022) Compositional and Metabolic Responses of Autotrophic Microbial Community to Salinity in Lacustrine Environments. *mSystems* 7: e00335–e22. <https://doi.org/10.1128/msystems.00335-22>
- Garritano AN, Song W, Thomas T (2022) Carbon fixation pathways across the bacterial and archaeal tree of life. *PNAS Nexus* 1: pgac226. <https://doi.org/10.1093/pnasnexus/pgac226>
- Goswami G, Hazarika DJ, Chowdhury N, Bora SS, Sarmah U, Naorem RS, Boro RC, Barooah M (2022) Proline confers acid stress tolerance to *Bacillus megaterium* G18. *Scientific Reports* 12: 8875. <https://doi.org/10.1038/s41598-022-12709-0>
- Gouffi K, Blanco C (2000) Is the accumulation of osmoprotectant the unique mechanism involved in bacterial osmoprotection? *International Journal of Food Microbiology* 55: 171–174. [https://doi.org/10.1016/S0168-1605\(00\)00192-6](https://doi.org/10.1016/S0168-1605(00)00192-6)
- Grossart H-P, Massana R, McMahon KD, Walsh DA (2020) Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. *Limnology and Oceanography* 65: S2–S20. <https://doi.org/10.1002/lno.11382>
- Herlemann DP, Geissinger O, Brune A (2007) The termite group I phylum is highly diverse and widespread in the environment. *Applied and Environmental Microbiology* 73: 6682–6685. <https://doi.org/10.1128/AEM.00712-07>
- Holmer M, Storkholm P (2001) Sulphate reduction and sulphur cycling in lake sediments: A review. *Freshwater Biology* 46: 431–451. <https://doi.org/10.1046/j.1365-2427.2001.00687.x>
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019) eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47: D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Hügler M, Sievert SM (2011) Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annual Review of Marine Science* 3: 261–289. <https://doi.org/10.1146/annurev-marine-120709-142712>
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11: 119. <https://doi.org/10.1186/1471-2105-11-119>
- Jiang X, Liu C, Cai J, Hu Y, Shao K, Tang X, Gong Y, Yao X, Xu Q, Gao G (2023) Relationships between environmental factors and N-cycling microbes reveal the indirect effect of further eutrophication on denitrification and DNRA in shallow lakes. *Water Research* 245: 120572. <https://doi.org/10.1016/j.watres.2023.120572>



- Karlicki M, Bednarska A, Hałakuc P, Maciszewski K, Karnkowska A (2024) Spatio-temporal changes of small protist and free-living bacterial communities in a temperate dimictic lake: Insights from metabarcoding and machine learning. *FEMS Microbiology Ecology* 100: fae104. <https://doi.org/10.1093/femsec/fae104>
- Kembel SW, Wu M, Eisen JA, Green JL (2012) Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Computational Biology* 8: e1002743. <https://doi.org/10.1371/journal.pcbi.1002743>
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Gloeckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41. <https://doi.org/10.1093/nar/gks808>
- Kraemer SA, Barbosa da Costa N, Shapiro BJ, Fradette M, Huot Y, Walsh DA (2020) A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. *The ISME Journal* 14: 3011–3023. <https://doi.org/10.1038/s41396-020-0733-0>
- Kuang B, Xiao R, Hu Y, Wang Y, Zhang L, Wei Z, Bai J, Zhang K, Acuña JJ, Jorquera MA, Pan W (2023) Metagenomics reveals biogeochemical processes carried out by sediment microbial communities in a shallow eutrophic freshwater lake. *Frontiers in Microbiology* 13. <https://doi.org/10.3389/fmicb.2022.1112669>
- Lefler FW, Berthold DE, Laughinghouse HD (2023) Cyanoseq: A database of cyanobacterial 16S rRNA gene sequences with curated taxonomy. *Journal of Phycology* 59: 470–480. <https://doi.org/10.1111/jpy.13335>
- Lew S, Glińska-Lewczuk K, Burandt P, Kulesza K, Kobus S, Obolewski K (2022) Salinity as a Determinant Structuring Microbial Communities in Coastal Lakes. *International Journal of Environmental Research and Public Health* 19: 4592. <https://doi.org/10.3390/ijerph19084592>
- Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* 22: 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)* 31: 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li H, Zeng J, Ren L, Wang J, Xing P, Wu QL (2017) Contrasting patterns of diversity of abundant and rare bacterioplankton in freshwater lakes along an elevation gradient. *Limnology and Oceanography* 62: 1570–1585. <https://doi.org/10.1002/lno.10518>
- Li H, Barber M, Lu J, Goel R (2020) Microbial community successions and their dynamic functions during harmful cyanobacterial blooms in a freshwater lake. *Water Research* 185: 116292. <https://doi.org/10.1016/j.watres.2020.116292>
- Li M, Li Q, Wang S, Wang X, Li Q, Liu W, Yu J, Zhang G, Wang J, Wu QL, Zeng J (2024) The diversity and biogeography of bacterial communities in lake sediments across different climate zones. *Environmental Research* 263: 120028. <https://doi.org/10.1016/j.envres.2024.120028>
- Lindström ES, Langenheder S (2012) Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports* 4: 1–9. <https://doi.org/10.1111/j.1758-2229.2011.00257.x>
- Linz AM, He S, Stevens SLR, Anantharaman K, Rohwer RR, Malmstrom RR, Bertilsson S, McMahon KD (2018) Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* 6: e6075. <https://doi.org/10.7717/peerj.6075>

- Liu Q, Yang J, Wang B, Liu W, Hua Z, Jiang H (2022) Influence of salinity on the diversity and composition of carbohydrate metabolism, nitrogen and sulfur cycling genes in lake surface sediments. *Frontiers in Microbiology* 13: 1019010. <https://doi.org/10.3389/fmicb.2022.1019010>
- Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, Rattei T, Damsté JSS, Spieck E, Le Paslier D, Daims H (2010) A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 107: 13479–13484. <https://doi.org/10.1073/pnas.1003860107>
- Margalef-Martí R, Sebilo M, Thibault De Chanvalon A, Anschutz P, Charbonnier C, Lauga B, Gonzalez-Alvarez I, Tessier E, Amouroux D (2023) Upside down sulphate dynamics in a saline inland lake. *Scientific Reports* 13: 3032. <https://doi.org/10.1038/s41598-022-27355-9>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12. <https://doi.org/10.14806/ej.17.1.200>
- McFadgen BG (1994) Archaeology and holocene sand dune stratigraphy on Chatham Island. *Journal of the Royal Society of New Zealand* 24: 17–44. <https://doi.org/10.1080/03014223.1994.9517454>
- McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8: e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Meredith AS, Croucher R (2007) State of the environment monitoring: water quality and ecosystem health of the lakes, streams and Te Whanga Chatham Island/Rehoku. *Environment Canterbury*.
- Nandy S, Kapley A (2024) Unraveling the potential of microbial communities for lake bioremediation via the metagenomics tool: A review. *AQUA – Water Infrastructure, Ecosystems and Society* 73: 11–33. <https://doi.org/10.2166/aqua.2024.154>
- Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M (2007) The vegan package. *Community ecology package* 10: 631–637.
- Pandey CB, Kumar U, Kaviraj M, Minick KJ, Mishra AK, Singh JS (2020) DNRA: A short-circuit in biological N-cycling to conserve nitrogen in terrestrial ecosystems. *The Science of the Total Environment* 738: 139710. <https://doi.org/10.1016/j.scitotenv.2020.139710>
- Pearman JK, Biessy L, Thomson-Laing G, Waters S, Vandergoes MJ, Howarth JD, Rees A, Moy C, Pochon X, Wood SA (2020) Local factors drive bacterial and microeukaryotic community composition in lake surface sediment collected across an altitudinal gradient. *FEMS Microbiology Ecology* 96. <https://doi.org/10.1093/femsec/fiaa070>
- Pearman JK, Adamson J, Thomson-Laing G, Thompson L, Waters S, Vandergoes MJ, Howarth JD, Wood SA (2022a) Deterministic processes drive national-scale patterns in lake surface sediment bacteria and eukaryotic assemblage composition. *Limnology and Oceanography* 68. <https://doi.org/10.1002/lno.12247>
- Pearman JK, Casas L, Michell C, Aldanondo N, Mojib N, Holtermann K, Georgakakis I, Curdia J, Carvalho S, Gusti A, Irigoien X (2022b) Comparative metagenomics of phytoplankton blooms after nutrient enrichment of oligotrophic marine waters. *Metabarcoding and Metagenomics* 6: e79208. <https://doi.org/10.3897/mbmg.6.79208>
- Peura S, Sinclair L, Bertilsson S, Eiler A (2015) Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. *Scientific Reports* 5: 12102. <https://doi.org/10.1038/srep12102>
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Gloeckner FO (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal

- RNA sequence data compatible with ARB. *Nucleic Acids Research* 35: 7188–7196. <https://doi.org/10.1093/nar/gkm864>
- R Core Team (2024) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rohwer RR, Kirkpatrick M, Garcia SL, Kellom M, McMahon KD, Baker BJ (2025) Two decades of bacterial ecology and evolution in a freshwater lake. *Nature Microbiology* 10: 246–257. <https://doi.org/10.1038/s41564-024-01888-3>
- Ruber J, Geist J, Hartmann M, Millard A, Raeder U, Zubkov M, Zwirgmaier K (2018) Spatio-temporal distribution pattern of the picocyanobacterium *Synechococcus* in lakes of different trophic states: A comparison of flow cytometry and sequencing approaches. *Hydrobiologia* 811: 77–92. <https://doi.org/10.1007/s10750-017-3368-z>
- Ruuskanen MO, St. Pierre KA, St. Louis VL, Aris-Brosou S, Poulain AJ (2018) Physico-chemical drivers of microbial community structure in sediments of Lake Hazen, Nunavut, Canada. *Frontiers in Microbiology* 9. <https://doi.org/10.3389/fmicb.2018.01138>
- Schallenberg LA, Pearman JK, Burns CW, Wood SA (2021a) Metabarcoding reveals lacustrine picocyanobacteria respond to environmental change through adaptive community structuring. *Frontiers in Microbiology* 12: 3258. <https://doi.org/10.3389/fmicb.2021.757929>
- Schallenberg LA, Pearman JK, Burns CW, Wood SA (2021b) Spatial abundance and distribution of picocyanobacterial communities in two contrasting lakes revealed using environmental DNA metabarcoding. *FEMS Microbiology Ecology* 97: fiab075. <https://doi.org/10.1093/femsec/fiab075>
- Song H, Li Z, Du B, Wang G, Ding Y (2012) Bacterial communities in sediments of the shallow Lake Dongping in China. *Journal of Applied Microbiology* 112: 79–89. <https://doi.org/10.1111/j.1365-2672.2011.05187.x>
- Tamames J, Puente-Sánchez F (2019) SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology* 9. <https://doi.org/10.3389/fmicb.2018.03349>
- Thomson-Laing G, Howarth JD, Vandergoes MJ, Wood SA (2022) Optimised protocol for the extraction of fish DNA from freshwater sediments. *Freshwater Biology* 67: 1584–1603. <https://doi.org/10.1111/fwb.13962>
- VanTassel NM, Janosik AM (2019) A compendium of Coastal Dune Lakes in Northwest Florida. *Journal of Coastal Conservation* 23: 385–416. <https://doi.org/10.1007/s11852-018-0671-4>
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73: 5261–5267. <https://doi.org/10.1128/AEM.00062-07>
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer. [https://doi.org/10.1007/978-3-319-24277-4\\_9](https://doi.org/10.1007/978-3-319-24277-4_9)
- Wood JM (2015) Bacterial responses to osmotic challenges. *The Journal of General Physiology* 145: 381–388. <https://doi.org/10.1085/jgp.201411296>
- Wu Z, Li M, Qu L, Zhang C, Xie W (2024) Metagenomic insights into microbial adaptation to the salinity gradient of a typical short residence-time estuary. *Microbiome* 12: 115. <https://doi.org/10.1186/s40168-024-01817-w>
- Yamamoto M, Arai H, Ishii M, Igarashi Y (2006) Role of two 2-oxoglutarate: ferredoxin oxidoreductases in *Hydrogenobacter thermophilus* under aerobic and anaerobic conditions. *FEMS Microbiology Letters* 263: 189–193. <https://doi.org/10.1111/j.1574-6968.2006.00415.x>

- Yau S, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, Rich J, Gibson JAE, Cavicchioli R (2013) Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *The ISME Journal* 7: 1944–1961. <https://doi.org/10.1038/ismej.2013.69>
- Yu G (2024) seqmagick: Sequence Manipulation Utilities. <https://cran.rstudio.com/web/packages/seqmagick/index.html> [July 17, 2024]
- Zhang J, Yang Y, Zhao L, Li Y, Xie S, Liu Y (2015) Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Applied Microbiology and Biotechnology* 99: 3291–3302. <https://doi.org/10.1007/s00253-014-6262-x>
- Zhou J, Ning D (2017) Stochastic community assembly: Does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews* 81: e00002–e00017. <https://doi.org/10.1128/MMBR.00002-17>

## Supplementary material 1

### Metadata for the lakes studied in the Chatham Islands alongside environmental and catchment data

Authors: John K. Pearman, Jack Sissons, Joseph Kanyi Kihika, Georgia Thomson-Laing, Susanna A. Wood

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.9.144128.suppl1>

## Supplementary material 2

### Indicative genes for specific pathways and the counts per million in each lake

Authors: John K. Pearman, Jack Sissons, Joseph Kanyi Kihika, Georgia Thomson-Laing, Susanna A. Wood

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.9.144128.suppl2>



### Supplementary material 3

#### Assembly information for the metagenomic samples

Authors: John K. Pearman, Jack Sissons, Joseph Kanyi Kihika, Georgia Thomson-Laing, Susanna A. Wood

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.9.144128.suppl3>

### Supplementary material 4

#### ASV table for the lakes with taxonomy

Authors: John K. Pearman, Jack Sissons, Joseph Kanyi Kihika, Georgia Thomson-Laing, Susanna A. Wood

Data type: csv

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.9.144128.suppl4>

### Supplementary material 5

#### Supplementary figure

Authors: John K. Pearman, Jack Sissons, Joseph Kanyi Kihika, Georgia Thomson-Laing, Susanna A. Wood

Data type: eps

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.9.144128.suppl5>